

NOTICE: This material may be protected by copyright law
(Title 71 U.S. Code).

Introduction

The aim of this book is to offer an introduction to the version of generative syntax usually referred to as Government and Binding Theory. I shall not dwell on this label here; its significance will become clear in later chapters of this book.

Government-Binding Theory is a natural development of earlier versions of generative grammar, initiated by Noam Chomsky some thirty years ago.¹ The purpose of this introductory chapter is not to provide a complete discussion of the history of the Chomskian tradition. A full discussion of the underpinnings of the generative enterprise would in itself be the basis for a book.² What I shall do here is offer a short and informal sketch of the essential motivation for the line of enquiry to be pursued. Throughout the book the initial points will become more concrete and more precise.

By means of footnotes I shall also direct the reader to further reading related to the matter at hand. Much of the primary literature will be hard to follow for the reader who has not worked his³ way through the book, but I hope that the information will be useful for future reference.

1 Linguistics: the Science of Language

When asked to indicate one prominent feature that distinguishes human beings from animals, many would probably say that this feature is 'language'. Even though animals may have communication systems, none of these systems is as rich or as versatile as the language used by humans. Language is human-

¹ For a survey of the development of the theory see van Riemsdijk and Williams (1986). This work should be accessible once chapter 7 has been covered.

² The reader will find a good introduction about generative grammar in general introductions to linguistics such as Akmajian, Demers and Harnish (1979), Fromkin and Rodman (1988), Lightfoot (1982), Smith and Wilson (1979), etc. These works should be accessible at this point. For more advanced introductions the reader is referred to Chomsky (1965, 1981a, b, c, 1982, 1986a, 1988), but reading them should be postponed until after chapter 7 of this book, at which point we shall have covered most of the technical issues that are discussed.

³ My use of the pronoun *his* for referents which may be either male and female follows the conventions of English grammar and I hope that the female readers of this book will not feel offended by it.

specific.⁴ This means that an understanding of the mechanisms of human language may lead us to understand, at least partly, what it is that distinguishes man from animals. Linguistics, the study of language, may give us an insight into the human mind.

Leonard Bloomfield defined linguistics as the science of language (Bloomfield, 1935). Like all scientists, linguists will aim at formulating the general rules and principles for the data with which they are faced. Linguists try to formulate generalizations about linguistic data, i.e. language.⁵

There are various ways of approaching the study of language. I assume the reader is familiar with the traditional view of language study, where the focus is often on the study of one specific language, say English. A linguist studying English will try to characterize the rules or principles that determine the formation of English sentences. The goal will be to provide a systematic description of English sentence formation, the grammar of English. The description will have to account for data such as the following:

- 1a Agatha Christie has written many books.
1b I don't like detective stories.

The sentences in (1) are *well formed*. They contrast with the sentences in (2), which are *ill formed*.

- 2a *Agatha Christie many books written has.
2b *I detective stories like.

Well formed sentences are constructed according to the rules of English grammar: they are *grammatical*. The sentences in (2) are not formed according to the rules of the grammar of English: they are *ungrammatical*, as indicated by the asterisks.

When writing a grammar, the linguist will not stop at merely listing examples with the appropriate grammaticality judgements. A simple catalogue of sentences may be an interesting basis for discussion but it will not satisfy the goal of scientific research. In addition to describing the data, the linguist will formulate general principles which will be applicable to further data. In-

⁴ In their introduction to linguistics Akmajian, Demers and Harnish (1979) present a fairly comprehensive discussion of the differences between human language and animal language.

⁵ Robins (1979) and Newmeyer (1980, 1983) offer good surveys of the development of linguistics. These books will offer a broader background to situate the theory we are discussing here in its historical context.

formal
instanc
hypoth
SVO p
SVO h
numbe
hypoth
it stan
object

3a Je
3b Jc
3c D
3d W

Eith
data i
the or
might
move
additi
you n
sente
Th
respe
a lan
the g
sente
A l
the l
string
sente
adeq

formally, a linguist might account for the ungrammaticality of (2), for instance, by proposing that in English verbs precede their direct objects. A first hypothesis might be that English sentences are constructed according to the SVO pattern: subject precedes verb, verb precedes object. Let us call this the SVO hypothesis. Having formulated this hypothesis on the basis of a limited number of data, the linguist will test it on the basis of further data. The SVO hypothesis will predict, for instance, that (3a) and (3b) are grammatical; but as it stands the hypothesis also predicts that (3c) and (3d) are ungrammatical: the objects, *detective stories* and *which stories* respectively, precede the subjects:

- 3a Jeeves is baking a cake.
3b John has bought a new car.
3c Detective stories, I don't like.
3d Which stories do you like?

Either the SVO hypothesis itself will have to be modified in the light of the data in (3c) and (3d) or one or more extra rules are needed which interact with the original hypothesis to account for the grammaticality of (3c) and (3d). We might, for example, formulate a rule of topicalization which allows one to move a direct object to the beginning of the sentence to account for (3c). In addition we might formulate a rule for question formation which states (i) that you move the questioning element (*which stories*) to the initial position of the sentence, and (ii) that you invert subject and auxiliary (*do*) (cf. (3d)).

The total of all the rules and principles that have been formulated with respect to a language constitutes the grammar of that language. A grammar of a language is a coherent system of rules and principles that are at the basis of the grammatical sentences of a language. We say that a grammar generates the sentences of a language.

A first requirement for any grammar is that it provides a characterization of the language it describes, i.e. the grammar must be able to distinguish those strings of words which are sentences of the language from those which are not sentences of the language in question. Such a grammar will be observationally adequate.

2 The Native Speaker: Grammaticality and Acceptability

2.1 Descriptive Adequacy

It is not only linguists that have the ability to judge English sentences. Every native speaker of English knows intuitively that the sentences in (1) and (3) are acceptable and that those in (2) are not. Moreover, every native speaker of English produces a large number of grammatical sentences and understands the English sentences that he comes across. The native speaker may not be able to state the rules and principles that underlie the sentences he produces, but he has an unconscious or tacit knowledge of such principles; he has internalized a grammar of the language.

The native speaker's tacit knowledge of the grammar of his language is the focus of enquiry for the linguist working in the Chomskian tradition. We say that a grammar reaches descriptive adequacy if it goes beyond the statement of rules to describe the facts and provides an account for the native speaker's intuitions.

Let us consider some examples. We have proposed that (3c) and (3d) could be generated by a rule that moves the direct object leftward to the beginning of the sentence. Now consider the examples in (4), which are not acceptable (hence the asterisk):

- 4a *Detective stories, I wonder if he likes.
 4b *Where do you wonder if he lives?

To account for the unacceptability of (4a) we might propose that the rule that moves the direct object in (3c) must be constrained: the direct object cannot move across *if*.

Similarly, when we consider (4b) we might propose that the rule of question formation must also be constrained: the questioning element (*where*) must not move across *if*. At this point we have reached observational adequacy: we provide an account for the facts. However, if we stop at this point we are missing a significant generalization. The ungrammaticality of (4a) and (4b) is due to the same constraint. A descriptively adequate grammar will not simply provide an analysis for (3c) and (3d) and for the deviance of (4a) and (4b), but it will try to capture the relation between (4a) and (4b) and formulate a general principle to explain why both (4a) and (4b) are felt to be unacceptable. Such a principle may be that no element in English must be moved across *if*. This

gene
whe

5a
5b

5c

6a
6b
6c

The
tak

lin
spe
acc
tac

lar
lin
of
co
gr

2.

At
'C
fo
'A
sp

7.
7.
7.
7
7

general rule will also predict that the examples in (5) are ungrammatical, whereas those in (6) are grammatical:

- 5a *Where do you wonder if Emsworth has hidden the Empress?
- 5b *Which detective do you wonder if Emsworth will invite for Sunday lunch?
- 5c *To Bill, I wonder if he will give any money.
- 6a Where has Emsworth hidden the Empress?
- 6b Which detective will Emsworth invite for Sunday lunch?
- 6c To Bill, he won't give any money.

The general constraint which blocks movement of an element across *if* will be taken to be part of the native speaker's internal grammar.

A descriptively adequate grammar will not only state rules to describe the linguistic data, but it will contain the general principles that enable the native speaker to produce and interpret sentences in his language and decide on the acceptability of sentences. Such a grammar is an explicit formulation of the tacit linguistic knowledge of the native speaker, his internal grammar.

The shift of focus from language itself to the native speaker's knowledge of language is a prime feature of the Chomskian tradition. Both the generative linguist and the traditional linguist will be constructing grammars, i.e. systems of rules that underlie the sentences of a language. But the generative linguist conceives of his grammar as a reflex of the native speaker's competence. The grammar is a representation of the speaker's internal linguistic knowledge.

2.2 Grammaticality and Acceptability

At this point we turn to the notions of 'grammaticality' and 'acceptability'. 'Grammaticality' is a theoretical notion. A sentence is grammatical if it is formed according to the grammar of English as formulated by the linguist. 'Acceptability', on the other hand, is the term which characterizes the native speaker's intuitions about the linguistic data. Consider (7):

- 7a Bill had left. It was clear.
- 7b [That Bill had left] was clear.
- 7c It was clear [that Bill had left].
- 7d Once that it was clear [that Bill had left], we gave up.
- 7e Once that [that Bill had left] was clear, we gave up.

(7a) contains two independent sentences. In (7b) the bracketed sentence *Bill had left* is the subject of the complex sentence *that Bill had left was clear*. We say that *Bill had left* is a subordinate clause. It is introduced by *that*, a subordinating conjunction. Similarly, in (7c) *that Bill had left* is a subordinate clause. In (7d) the sentence (7c) is a subordinate clause in a complex sentence. A grammar must include a rule to generate complex sentences in which one clause is part of another one.

Let us turn to (7e). The sentence is odd for most native speakers: it is not acceptable. However, this sentence is formed according to the same principle that we posited to account for the formation of (7b)–(7d), i.e. that one sentence may become part of another sentence. Hence (7e) would be **grammatical**: it is formed according to the rules of the grammar, though it is not acceptable.

Faced with intuitions such as that for (7e) the linguist might decide to modify the rules of the grammar he has formulated in such a way that sentence (7e) is considered to be ungrammatical. He may also decide, however, that (7e) is grammatical, and that the unacceptability of the sentence is due to independent reasons. For instance, (7e) may be argued to be unacceptable because the sentence is hard to process. In the latter case the unacceptability is not strictly due to linguistic factors but is due to the more general mechanisms used for processing information.

The native speaker who judges sentences cannot decide whether the sentence is formed according to the rules of the grammar. He only has intuitions about acceptability. It is for the linguist to determine whether the unacceptability of a sentence is due to grammatical principles or whether it may be due to other factors. It is the linguist's task to determine what it is that makes (7e) unacceptable.

This entails that there may be disagreement between linguists as to whether certain unacceptable sentences are grammatical or not. The disagreement is not one of conflicting judgements of the sentence (although these may also exist), but it is one of analysis. The linguist will have to determine to what degree the unacceptability of a sentence is to be accounted for in terms of the grammar. All the linguist has to go by, though, is the native speaker's intuitions about language, and these, as argued above, are the result of the interaction between his internal grammar and other factors.

In this book we focus on the linguistic knowledge of the native speaker. We restrict our attention to his internal grammar. Obviously, the interaction between the grammar and other mental processes is also an interesting area of research, but it is not the topic of this book.

2.3 The Grammar as a System of Principles and Rules

One approach to formulating a grammar of a language would be to suppose that the speaker's internal knowledge of English, i.e. his internal grammar, is no more than a huge check-list of grammatical sentences. Speakers could be thought to 'check' any sentence they come across against this internal inventory. Sentences which match a sentence in the list would be said to be grammatical, those that do not are ungrammatical. Depending on the degree of deviance of such ungrammatical sentences we could rank the sentences for ungrammaticality. A grammar of a language would then be simply a list of sentences.

But it must be immediately obvious that listing all the grammatical sentences of a language is an impossible task and also that it misses the point.

Cataloguing all the grammatical sentences of English is first of all impossible because there is an infinite number of English sentences. In addition, there are other objections to such a listing enterprise. We stated above that linguistics is the scientific study of language. From such a perspective the listing of linguistic data is not enough. We expect general principles to explain the data.

For the generative linguist who tries to provide a representation of the native speaker's internal knowledge of a language a mere listing of sentences would never achieve descriptive adequacy: it could never account for the native speaker's knowledge of the language. Human beings – in our example speakers of English – have finite memories: we often forget things we have heard (perhaps unfortunately). Given that the capacity of our memories is finite, it would be absurd to claim that human beings are able to store all potential sentences of the language, an infinite set. It is thus inconceivable that the native speaker's internal linguistic knowledge consists in an inventory of sentences. We must assume that human beings are somehow equipped with a finite system of knowledge which enables them to construct and interpret an infinite number of sentences. This finite system of principles is what we referred to loosely above as the internal grammar of the language. The generative linguist will try to render explicit the finite system of rules and principles that make up the native speaker's competence. In our example, the rule which prohibits moving elements across *if* will be able to account for the unacceptability of (4) and (5).

3 Knowledge of Language: Universal and Specific Properties of Language

3.1 Explanatory Adequacy and Language Acquisition

Suppose that we have achieved our goal and that we have provided an explicit characterization of the system of general rules and principles of English sentence formation and which we assume is a representation of the native speaker's tacit knowledge of his language, in our example English. Our grammar will then have reached both observational and descriptive adequacy. Now another important and indeed fascinating question arises. We would like to determine how native speakers of a language, in our example English, come to possess the knowledge of their language. We shall say that a theory reaches **explanatory adequacy** if it can account for the fact that the principles of the internal grammar can get to be known to speakers.

This question is obviously intriguing since millions of people are native speakers of English without being able to formulate any of the rules of the grammar of their language. And all these people had learnt English more or less by the age of six. On the other hand, millions of people do not speak English at all but are fluent speakers of French, Italian, Japanese or any of the other languages that exist.

It would not be reasonable to argue that speakers of English are taught everything they know about their language. Native speakers are not formally taught, for instance, to judge the sentences in (4) and (5) above, and yet they will all agree on their unacceptability. How then do they attain this knowledge? To illustrate the same point, let us consider the following examples, taken from Chomsky (1986a: 8):

- 8a I wonder who [the men expected to see them]
 8b [The men expected to see them]

Both (8a) and (8b) contain the string *the men expected to see them*. In (8a) the pronoun *them* can be interpreted as referring to *the men*; in (8b) it cannot. The difference in interpretation of the pronoun is not something that is formally taught and there is no overt indication in the sentences to signal the difference. Still, all native speakers of English have the intuition; non-native speakers of English too will never interpret the pronoun *them* in (8b) as referring to *the men* even though this is not part of their formal tuition in English.

Give
of the
speake
idea is
langua
beings
unreas
speake
others
born
beings
innate
just on

3.2

Let u
introc
(7) h
senter
comp

9

The c
the n
Engli
is pa
embe
Fren

10a

6

Given that neither formal teaching nor overt evidence seems to be the source of the native speaker's intuitions, it is proposed that a large part of the native speaker's knowledge of his language, i.e. the internal grammar, is innate. The idea is that human beings have a genetic endowment that enables them to learn language. It is this innate capacity for language learning common to all human beings that the generative linguist tries to characterize. Of course, it would be unreasonable to posit that some individuals – those that will become native speakers of English – are born with a specific grammar of English and that others – those that will end up speaking Japanese as their first language – are born with the grammar of Japanese readily stored in their minds. Human beings with normal mental faculties are able to learn *any* human language. The innate linguistic endowment must be geared to any human language and not to just one.

3.2 Universal Grammar

Let us discuss some examples to try to clarify all this a little. We have introduced one generalization about English: the SVO hypothesis. The data in (7) have led us to formulate another hypothesis: any grammatical English sentence can apparently be embedded and become a subordinate clause in a complex sentence. Let us refer to this as the embedding principle.

9 Embedding principle⁶

A grammatical sentence can become a subordinate clause in a complex sentence.

The embedding principle tries to render explicit part of the tacit knowledge of the native speaker. This principle would be taken to be part of the grammar of English, hence available to the native speaker. But this principle is not one that is particular to the grammar of English, it is not *language-specific*. Rather, the embedding principle is part of the grammar of all human languages. Thus in French too we find sentences such as (10a) embedded in (10b):

- 10a Maigret a abandonné l'enquête.
Maigret has abandoned the enquiry.

⁶ As the reader will see later, the embedding principle is not in fact part of our grammar. The fact that sentences can be embedded can be deduced from the principles of sentence formation discussed in chapters 1 and 2.

- 10b Lucas a annoncé que Maigret a abandonné l'enquête.
Lucas has announced that . . .

The reader who knows other foreign languages will be able to apply this principle to those languages.

The embedding principle is a **universal**⁷ principle. Principles that hold of all languages are said to be part of **universal grammar**, or UG for short. Informally, UG is a system of all those principles and rules that are common to all human languages, this means languages as different as English and French or Japanese.

A hypothesis adopted by generativists of the Chomskian tradition is precisely that universal grammar is innate to the human species. UG is a genetic endowment: we are born equipped with a set of universal linguistic principles. To quote Chomsky himself: 'Universal grammar may be thought of as some system of principles, common to the species and available to each individual prior to experience' (1981b: 7).

If we assume that there is such an innate linguistic endowment the task of attaining the knowledge of a specific grammar, say English, is facilitated. Someone learning English would not have to learn the embedding principle. It is innate; it is part of one's genetic endowment.⁸

Universal grammar is the basis for acquiring language. It underlies all human languages. All and only human beings are equipped with UG and they are all able to learn languages. Other systems (say, dogs or television sets) are not equipped with UG and therefore will not be able to learn human languages. The linguistic endowment characterized as UG is species-specific.

3.3 Parameters and Universal Grammar

The innate linguistic endowment UG is not sufficient to enable one to speak a language. If all we needed was UG then human beings would be able to speak any language wherever they were born and in whatever circumstances they grew up. The native language is that spoken by the child's immediate environment. It would be inconceivable, for instance, that a child growing up

⁷ The principle might have to be reformulated and reinterpreted in the light of further data and in view of other hypotheses of our linguistic theory (for an example see the discussion of word order patterns in German in chapter 11).

⁸ The reader may wonder why, if the principle is innate, children do not start using complex sentences straight away. However, it is conceivable that the development of the internal grammar interacts with a general maturation process. We leave this problem aside here.

in a com
Japanese
and the
well-kn
in adult
While
lot of ve
importa
gramma
Japanes
English
object p

11 Jo!
Jo!
(K

Engl:
as subj
ordered
cannot
(and o
How c
followi
other t
univers

Subj
the chi
his lan
let us
differ
what t
learnin
param
section

The
task h
the ch
his li
Sente
subje

in a community where only English is spoken could become a native speaker of Japanese. Human beings usually master one language with native competence and they have a hard time learning other languages later in life. It is a well-known fact that achieving complete mastery of second or third languages in adulthood is exceptional.

While certain grammatical principles and rules are universal, there is also a lot of variation between different languages. The grammar of English differs in important respects from that of, say, Japanese. Hence, if you 'know' the grammar of English, this will not entail that you 'know' the grammar of Japanese. In (1) we illustrated some simple English sentences and we saw that English sentences exhibit SVO word-order. In Japanese, on the other hand, the object precedes the verb; Japanese is SOV:

- 11 John ga Mary o but-ta.
 John particle Mary particle hit-past
 (Kuno, 1973: 3)

English and Japanese are similar in that sentences contain components such as subjects, objects and verbs. But they differ in the way these elements are ordered. The SVO hypotheses which we postulated as part of English grammar cannot be an absolute linguistic universal: it is part of the grammar of English (and of other languages) but not of that of Japanese. It is language-specific. How does the child learning English attain this rule? We could envisage the following scenario. The linguistic endowment UG makes available, among other things, the notions 'subject', 'object', 'verb'. Let us propose that these are universal concepts, available in all human languages.

Subject, verb and object will be linearly ordered. When learning a language the child will have to decide which is the word-order pattern characteristic of his language. We say that there exists a **parameter** along which languages vary: let us call it the word-order parameter. Individual languages will illustrate different 'settings' or 'values' of the parameter. The child's task is to figure out what the particular value of the parameter in question is for the language he is learning. The different word-orders of English and Japanese are the result of **parametric variation**. Another example of parametric variation is discussed in section 4.

The child must construct his internal grammar of English. To achieve this task he uses, on the one hand, the universal notions and principles of UG and the choices that it makes available, and on the other hand he uses the data of his linguistic experience, in our example the English sentences he hears. Sentences such as those in (1) will provide evidence to the child that in English subject precedes verb and verb precedes object. A sentence such as that in (11)

will enable the child exposed to Japanese data to decide that Japanese is SOV.

Exposure to linguistic material is an essential ingredient in the child's learning process. The child will need the linguistic experience to start constructing the internal grammar of his language and thus to attain the knowledge of a language. Without exposure the child would not be able to construct his internal grammar: he would not know, for instance, whether to pick SVO or SOV. UG is crucial in the organization of the primary linguistic experience. UG determines the way the child will interpret and organize the language he is exposed to. In our example, UG would be said to make available the concepts subject, object and verb, and the word-order parameter, and the child will fix the setting of the parameter. We have now postulated two properties of UG:

- (i) UG contains a set of absolute universals, notions and principles which do not vary from one language to the next.
- (ii) There are language-specific properties which are not fully determined by UG but which vary cross-linguistically. For these properties a range of choices is offered by UG. One parameter along which languages vary concerns word-order.

We have already said that absolute universal principles need not be learnt. But even with respect to the mastery of language-specific properties very little 'learning' is involved under the hypothesis outlined above. For those principles that are parametrized, the available options are given by UG. Attaining linguistic knowledge consists in fixing the parameters.

From this point of view, we conclude that the mastery of a language is not really the result of learning. Rather, being equipped with UG (with its parameters) and exposed to a language, the child cannot but construct the grammar of the language he is exposed to. For this reason the term 'learning' is often replaced by the term 'acquisition'.

In addition, the exposure to language will also equip us with a vocabulary, the words of the language to which we are exposed. Even if we have an innate knowledge of the principles of language we must inevitably learn the lexicon of the language, the words and their meaning, in order to be able to put this knowledge into operation. Thus an English child will have to learn all the words in the sentences above, and indeed many more. And we go on learning new words throughout our lives. Similarly a French child will learn the French lexicon, etc.⁹

⁹ The acquisition of the vocabulary of a language is also a matter of interest. For some introductory discussion the reader is referred to Lightfoot (1982: 121-2).

To s
uncons
langua
partic
proces
Exp
The vi
curren

3.4

Our a
param
a spec
of one
Sch
as foll

12

The
princi
the la
also l

I
r
i
C
f
t
a

10
I
C
V
i

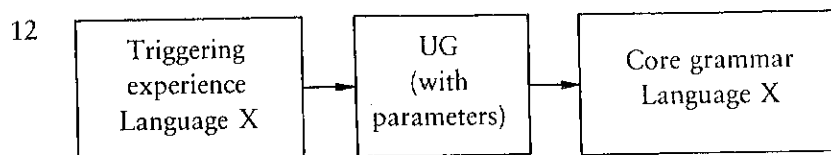
To sum up so far: human beings are born equipped with some internal unconscious knowledge of grammar: UG. UG is a set of universal principles of language, some of which parametrized. Via the input of the experience of one particular language this knowledge can be implemented. The acquisition process is 'triggered' by the exposure, the child's linguistic experience.

Exposure will also enable the child to learn the vocabulary of the language.¹⁰ The view of language acquisition in terms of parameter setting is the basis of current work in the generative tradition.

3.4 Language Learning and Language Acquisition

Our ability to speak a language is based partly on the innate principles and parameters available in UG, partly on the triggering experience of exposure to a specific language. On the basis of these components we develop a grammar of one (or more) specific languages: the 'core grammar' of such a language.

Schematically we can represent the generative view of language acquisition as follows:



The exposure to some language, say English, will activate the innate principles of universal grammar. The child will fix the choices to be made for the language in question, for instance, that the object follows the verb, and will also learn the vocabulary of the language. To quote Chomsky:

Endowed with these principles, a system provided with adequate experience will develop a grammar of the peculiar and specific sort characteristic of human language . . . Lacking these principles, a system will develop no grammar or some different system. The telephone exchange, for example, has 'heard' much more English than any of us, but lacking the principles of universal grammar . . . it develops no grammar of English as part of its internal structure. (1981b: 8)

¹⁰ The reader will find a very accessible discussion of the acquisition process in Lightfoot (1981, 1982). For further information the reader should consult Chomsky (1981a, b and c) and the literature cited there. However, Chomsky's work will be hard to read at this stage and the reader is advised to postpone reading these works until he has worked through chapters 1-7 of this book.

By the age of six a child exposed to English will have constructed the grammar of his language. This does not mean that no further development of his knowledge of language is possible. For instance, we go on learning new words throughout our lives. In addition we also learn certain less usual constructions of the language. These exceptional or marked patterns of the language are not taken to be part of the *core grammar* of the language, they belong to the *marked periphery* of the grammar and may be acquired later. The native speaker will also have to learn all of the social or cultural conventions associated with his language, for instance, that certain words belong to a very high style whereas others are informal. These conventions are not part of the grammar, they belong to the more general domain of human behaviour.

The aim of generative syntacticians is to develop a theory of language that is a model of the acquisition of language. Linguists want to provide an explicit formulation of the three components of (12): (i) the principles of UG and the parametric variation across languages, (ii) the triggering experience needed to activate the principles of UG, and (iii) the core grammar of specific languages as it derives from these interacting components. A theory that can account for these three components will be said to have reached explanatory adequacy.

4 The Generative Linguist

The research programme as sketched here briefly and roughly is one that has been motivating linguistic research for the past thirty years and has given rise to many challenging results. The programme is indeed still developing.

It may be useful to repeat that the ultimate aim of generative linguistic theory is not to describe the details of one specific language, but rather to formulate the underlying principles that determine the grammars of human languages. These grammars are seen as representations of the native speaker's knowledge. In the course of their enquiry, linguists will examine data drawn from individual languages, of course, but the investigator will always bear in mind the interacting components in (12).

The generative linguist who tries to characterize knowledge of a language, say English, will wish to do two things: (i) he needs to determine what properties of English are universal, and (ii) what properties are English-specific and how these relate to the parameters of UG.

It must by now have become clear that by simply looking at English and only that, the generative linguist cannot hope to achieve his goal. All he can do is

write a
but he
speake
with o
are un
by un
gramm
engage
Wo
Gener
use an
choice
impor
Both l

5 I

Let us
(12) a
acquis
any o
leave
Ho
child
rules
We
will le
be ex
subje

13a

13b

13c

write a grammar of English that is observationally and descriptively adequate but he will not be able to provide a model of the knowledge of the native speaker and how it is attained. The generativist will have to compare English with other languages to discover to what extent the properties he has identified are universal and to what extent they are language-specific choices determined by universal grammar. Even when his main concern is some aspect of the grammar of English the linguist will have to go outside this one language and engage in contrastive work.

Work in generative linguistics is therefore by definition comparative. Generative linguists often do not focus on individual languages at all: they will use *any* human language to determine the general properties of UG and the choices it allows. Data from a language spoken by millions of people are just as important as data from a dialect spoken by only a couple of hundred people. Both languages are human languages and are learnt in the same way.

5 Language Acquisition: Some Speculation

Let us try to think a little more about language acquisition as represented in (12) above. A first proviso is in place: we are only looking at language acquisition from a linguistic point of view. We abstract completely away from any other psychological aspects involved in language acquisition. We also leave aside the mastery of social conventions concerning language use.

How does a child construct the grammar of his language? By hypothesis the child comes to the task equipped with UG: a set of universal principles and rules some of which are parametrized.

We illustrate these parameters with another example. Speakers of Italian will learn that in their language the pronominal subject of a sentence need not be expressed, while for French and English, speakers need to learn that the subject of a sentence is expressed, even if it is pronominal:

- 13a Ho incontrato il commissario Maigret.
 Io ho incontrato il commissario Maigret.
 I have met the inspector Maigret
- 13b *Ai rencontré le commissaire Maigret.
 OK J'ai rencontré le commissaire Maigret.
 I have met the inspector Maigret
- 13c *Have met inspector Maigret.

In (13a) we see that in Italian the subject pronoun *io* ('I') may be left unexpressed. In (13b), the French example, the sentence is ungrammatical if the subject pronoun is not expressed. In English too (13c) will be ungrammatical in normal conversation.¹¹

Whether the subject of a sentence must be expressed overtly or not is subject to parametric variation. Learners of specific languages have to fix the choice, the setting of the parameter. Let us call it the **null subject parameter**. The null subject parameter is positively set in Italian. It is negatively set in English and French.¹²

14a Null subject parameter

Subject pronouns may be dropped.

14b Parameter settings: null subject parameter:

- French —
- English —
- Italian +
- etc.

Fixing the parameter is easy enough for the Italian child. Data such as sentence (13a) will trigger the positive setting for the parameter. But how can an English child who hears only examples with overt subjects deduce that sentences such as (13c) are impossible? After all, it might be a mere accident that he is only faced with sentences with full pronouns? The problem described here is that of **negative evidence**: how do you conclude that something is not possible merely on the basis that you have not met with any occurrences of it?

One proposal to solve this problem is as follows. As a starting point observe that with respect to the realization of sentences with pronouns as subjects, Italian is a 'bigger' language than English. For every English sentence with a pronoun as its subject, there is the additional option in Italian of leaving the pronoun unexpressed:

¹¹ In certain areas of usage the subject may be non-overt in English too:

- (i) Wish you were here.
- (ii) Had a nice day at the swimming pool.

(i) is typical of informal writing, (ii) could be a sentence taken from a diary. These sentences are acceptable within certain registers but not generally. We shall say that they are not part of the **core grammar** of English but belong to the marked **periphery**. For some discussion see Haegeman (1990).

¹² Chapter 8 of this book contains a discussion of this parameter.

15 Eng

16 Ital

The se
in Italian

17

Some
and fixin
assumpt
'smalles
start ou
produce

The ic
bigger l
subjects
be given
subject
the para

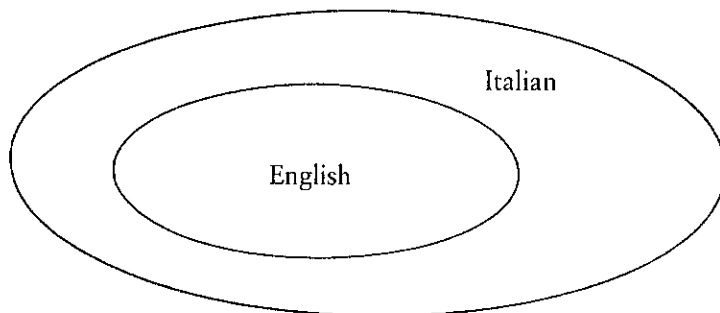
Work
mentior
research

13 An
sub
atte
acq
198

- 15 *English*: She is ill.
 *Is ill.
16 *Italian*: Lei è malata.
 È malata.

The sentences with pronoun subjects in English are a subset of those possible in Italian. For each sentence in English there are two Italian equivalents:

17



Some linguists have suggested that when constructing his internal grammar and fixing the values of the parameters, the child starts out with the minimal assumption. He starts from that parameter setting which produces the 'smallest' language. In our concrete example this would mean that all children start out assuming that the null subject parameter is negatively set, which will produce a language in which sentences without subjects are excluded.

The idea then is that the child will only opt for the setting which produces a bigger language, in our case a language in which sentences without visible subjects are allowed, if he is given positive evidence. The English child will not be given evidence to suggest that his language is positively specified for the null subject parameter; the Italian child will. Hence only the Italian child will reset the parameter.¹³

Work in this direction is very much in progress but it is important to mention it because it ties in directly with the general motivation of generative research in language. Recently, linguists in the generative tradition have also

¹³ An account in which language acquisition is related to the notions of sets and subsets is developed by Manzini and Wexler (1987). The reader should not attempt to read this text until he has studied chapter 7. For a different view on the acquisition of the null subject parameter the reader is referred to Hyams (1986, 1989).

started to investigate whether the model of first language acquisition that they advocate could be applied to the acquisition of a second language.¹⁴

Introduction and Overview

In the Introduction we saw that a grammar of a language is a coherent system of rules and principles which determines the formation of the sentences of a language. The basic unit with which a grammar is concerned is the sentence. A grammar will specify what the components of the sentence are, how they interact, in which order they occur, etc. Partly, the rules formulated will be of a universal nature; partly, they will have to be parametrized to bring out language-specific properties of individual languages.

Grammars have nothing to say about units higher than the sentence, such as the paragraph, the discourse exchange, the text, etc. Such higher units will be the object of another type of enquiry.¹

In this chapter we consider the relation between the structure of the sentence and the words that make up the sentence. We shall see that sentence structure is to a large extent determined by lexical information. As pointed out in the Introduction, it is assumed that the reader is familiar with the basic techniques and terminology of sentence parsing.

Chapter 1 is organized as follows: section 1 provides a brief discussion of the central concepts of sentence structure; section 2 focuses on the relation between lexical items and sentence structure; section 3 discusses the predicate-argument structure of sentences and introduces theta theory; section 4 sums up the link between lexical items and sentence structure and introduces the projection principle; section 5 explores the application of theta theory, concentrating on clausal arguments, expletive (non-argument) pronouns and auxiliary verbs; section 6 discusses the general constraint that sentences must have subjects; and in section 7 we consider the properties of the subject theta role.

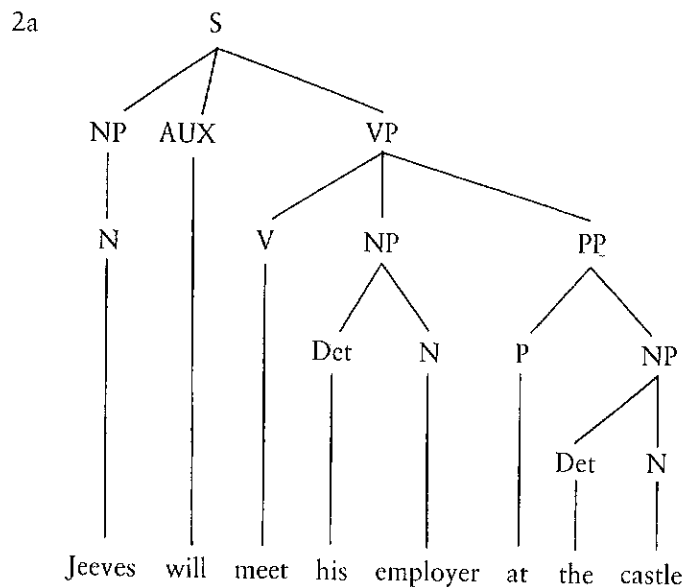
1 The Units of Syntactic Analysis

In this section we briefly recapitulate the basic notions of syntactic structure that will be the starting point for our discussion. Consider the following example:

¹ For an interesting approach to the study of sentences in discourse see Sperber and Wilson (1986) and Kempson (1988), who examines the link between Sperber and Wilson's theory of utterance interpretation and formal syntax.

1 Jeeves will meet his employer at the castle.

(1) is a grammatical English sentence. When we look for its component parts, the constituents, the units that perhaps come to mind first are the words of the sentence: sentence (1) contains eight words. But, as anyone familiar with traditional techniques of sentence parsing knows, words are not the **immediate constituents** of a sentence. Rather, they are the **ultimate constituents**. The words of the sentence are organized hierarchically into bigger units called **phrases**. In the framework of generative syntax the constituent structure of a sentence is represented in one of the following formats: by means of the tree diagram format as in (2a), by means of phrase structure rules or rewrite rules as in (2b), or by means of labelled brackets as in (2c).²



- 2b
- i S → NP – AUX – VP
 - ii NP → (Det) – N
 - iii VP → V – NP – PP
 - iv PP → P – NP
 - v N → *Jeeves, employer, castle*
 - vi V → *meet*

² For an introduction to parsing see Burton-Roberts (1986), Fromkin and Rodman (1988) and Wekker and Haegeman (1985).

- vii AUX \longrightarrow *will*
- viii P \longrightarrow *at*
- ix Det \longrightarrow *the, his*

- 2c [S [NP [N Jeeves]] [AUX will] [VP [V meet] [NP [Det his] [N employer]]
[PP [P at] [NP [Det the] [N castle]]]]]

Representations such as those in (2) give us information concerning the structure of (1). They indicate, for instance, that the string *his employer* is a syntactic unit, a **constituent**. It is a **noun phrase (NP)**, a constituent whose main element or **head** is the noun (N) *employer*. Analogously, the constituent *at the castle* is a **prepositional phrase (PP)**; the head of this PP is the preposition *at*, which is followed by an NP, *the castle*. The constituent *meet his employer at the castle* is a **verb phrase (VP)**, whose head is the verb *meet*, which is followed by two NPs: the NP *his employer* and the PP *at the castle*.

The structural representations in (2) allow us also to describe syntactic operations that may affect sentence (1). Consider (3):

- 3a At the castle, Jeeves will meet his employer.
- 3b His employer, Jeeves will meet at the castle.
- 3c Meet his employer at the castle, Jeeves will (indeed).

The sentences in (3) are intuitively felt to be variations upon sentence (1); they are all paraphrases of (1). In order to capture the similarity between the sentences in (3) and that in (1) we shall assume that all these sentences have the same **underlying structure**, represented in (2). In each of the sentences in (3) one of the constituents identified in (2) has been moved to the beginning of the sentence, or **preposed**. Thus in (3a) the PP *at the castle* has been moved, in (3b) the NP *his employer* is preposed, in (3c) the VP, *meet his employer at the castle*, is preposed. The possibilities for preposing elements of a sentence can be seen to be structure-based. Only constituents of the sentence such as **NP** and **VP** can be preposed. One cannot indiscriminately prepose any random string of words in the sentence:³

- 3d *Employer at the, Jeeves will meet his castle.
- 3e *Meet his, Jeeves will employer at the castle.

³ For a formal discussion of operations such as preposing, see chapters 6 and 7.

nt parts,
ds of the
ar with
mediate
ts. The
s called
ure of a
the tree
ite rules

Another operation that affects sentence constituents is the one that forms questions. If we form questions on the basis of (3) we see that again the constituent structure represented in (2) plays a crucial role.

We distinguish two types of questions: *yes-no* questions and constituent questions. The classification adopted is based on the type of answer expected. (4a) is a *yes-no* question: in normal circumstances we expect *yes* or *no* as an answer. The other questions in (4) are constituent questions: the answer to the question will be a constituent.

- 4a Will Jeeves meet his employer at the castle?
- 4b Who will Jeeves meet at the castle?
- 4c Where will Jeeves meet his employer?
- 4d What will Jeeves do?
- 4e Who will meet his employer at the castle?

Yes-no questions are formed by moving the auxiliary (here, *will*) to sentence-initial position. Constituent questions are formed by means of a substitution process. Each of the sentence-initial question-words such as *who*, *where*, *what* in (4) substitutes for one of the constituents identified in (2): in (4b) *who* substitutes for the object NP *his employer*.⁴

Operations such as preposing and question formation thus provide evidence for the role of phrase structure in syntax.

2 Words and Phrases

Although words are not the immediate constituents of the sentence, they play an important role as the ultimate building blocks of the sentence.

Words belong to different **syntactic categories**, such as nouns, verbs, etc., and the syntactic category to which a word belongs determines its **distribution**, that is, in what contexts it can occur. Normally, one cannot easily interchange words of one category for words of another. If you were to replace the verb *meet* by the semantically-related noun *appointment* in (1) you would no longer obtain a grammatical sentence:

- 5 *Jeeves will appointment his employer at the castle.

⁴ Chapter 7 contains a detailed discussion of the formation of questions.

The grammar of English, and indeed of any language, will have to have access to the categorial information attached to lexical items since this information plays a part in the formation of sentences.

We assume that the categorial information is also available to the native speakers of the language: they will agree that (5) is unacceptable and that the unacceptability is due to the inappropriate use of the N *appointment*. We postulate that speakers of a language are equipped with an internal 'dictionary', which we shall refer to as the mental lexicon, or lexicon, which contains all the information they have internalized concerning the words of their language. As seen above, this mental lexicon will have to contain, among other things, information on syntactic categories. We assume that each word of the language known by a speaker will be listed in his mental lexicon with its categorial specification. For instance, a native speaker of English will presumably have a lexicon containing the following information:

6a	<i>meet</i> :	verb
6b	<i>employer</i> :	noun
6c	<i>castle</i> :	noun
6d	<i>at</i> :	preposition
6e	<i>the</i> :	determiner
6f	<i>his</i> :	determiner
6g	<i>appointment</i> :	noun

As we suggested in the Introduction, it would not make sense to claim that the native speaker's lexical knowledge, i.e. the mental lexicon, is entirely innate. If lexical knowledge were completely innate, then human beings would have to be born equipped with the lexicons of all known or possible human languages. Rather, we assume that the lexicon of a language is learnt by each native speaker. The speaker learns the words of the language and what category they belong to. But this does not imply that he comes to this learning process totally unprepared. We assume that UG, our innate knowledge of language, contains, for example, the notion of syntactic category. When exposed to the words of a particular language, speakers will have some expectation as to which categories to discover. We shall not speculate further here as to the sort of knowledge this involves.

Lexical information plays a role in sentence structure because the syntactic category of a word determines its distribution. Let us take as an example sentence (1) and consider its syntactic representation (2a). In the tree diagram (2a) the word *appointment* will not be inserted in a position dominated by the node V because only verbs can be inserted under a node V, the same observation would apply to the other words in the sentence. Looking at the

tree diagram from top to bottom we can say that the terminal category labels such as N, V, etc. restrict which lexical elements can be inserted.

Looking at the tree from bottom to top, we see that the words that are inserted at the bottom of the tree determine the structure of the sentence. The inserted words will determine the syntactic category of the head of the phrase and hence they will ultimately determine the category of a phrase, the **phrasal category**. For instance, in our example (2a) the inserted N *employer* will be the head of a phrase of the type NP and not of a VP. Chapter 2 provides a more detailed discussion of the principles that regulate sentence structure.

Clearly, the mere matching of lexical and phrasal categories is not sufficient to produce a good sentence. For instance, the random insertion of nouns in the slots provided for them in (2) produces odd results in (7b) and (7c):

7a Jeeves will meet his employer at the castle.

7b ? Jeeves will meet his castle at the meeting.

7c ? Jeeves will meet his castle at the employer.

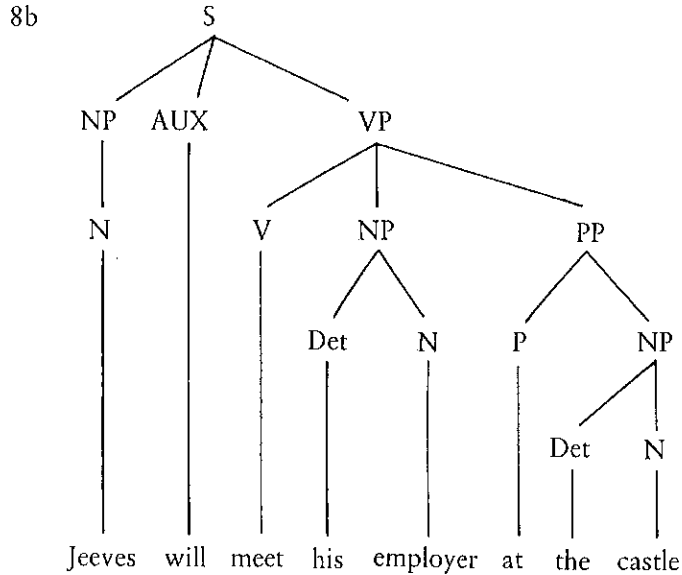
The question arises whether (7b) and (7c) are ungrammatical: is their oddness due to a violation of a grammatical rule or principle? When asked to explain what is displeasing in (7b), a native speaker will say that (7b) is bizarre because the verb *meet* is followed by the string *his castle*. The oddness is due to the fact that the concept 'meet' usually involves an interaction between two animate participants, while 'his castle' refers to an inanimate entity which does not normally qualify to take part in an action of the type 'meet'. But if we were to endow the concept 'castle' with animacy the oddness would be removed. In a fairy tale where castles take a walk (7b) would become acceptable. What is wrong with (7b) is not a grammatical issue; its strangeness relates to our general knowledge of the world. Issues of language use which hinge on the interaction of the grammar with extra-linguistic information such as that just described must not be integrated in a grammatical description. Grammars do not contain principles and rules about our beliefs about the world around us. (7b) may therefore be seen as grammatical but as bizarre in view of our encyclopaedic knowledge of castles as inanimate objects.

Let us return to sentence (1), repeated here as (8a), and its tree diagram representation (2a) repeated here as (8b):

8a Jeeves will meet his employer at the castle.

8b

It is
esse
plac
can8c
8dIt
cert3
—In
req
for



It is clear that some of the phrasal constituents of this sentence are more essential to the sentence than others. The PP *at the castle*, which specifies the place of the event, can be left out without any harm, but the NP *his employer* cannot.

8c Jeeves will meet his employer.

8d *Jeeves will meet at the castle.

In the next section we shall try to account for the obligatory nature of certain constituents in a sentence.

3 Predicates and Arguments

In this section we explain which constituents of a sentence are minimally required, and why. We first provide an informal discussion and then develop a formal approach known as theta theory.

3.1 Subcategorization

Consider the following sentences:

- 9a Maigret will [_{VP} imitate [_{NP} Poirot] [_{PP} with enthusiasm]].
 9b Bertie will [_{VP} abandon [_{NP} the race] [_{PP} after the first lap]].
 9c Miss Marple will [_{VP} reconstruct [_{NP} the crime] [_{PP} in the kitchen]].

The labelled bracketing in (9a), (9b) and (9c) shows that these sentences are structurally similar to sentence (8a) with its representation in (8b). In (8a) and in each of the sentences in (9) the VP contains a V, the head of the VP, an NP, the direct object, and a PP. In each of these the PP is optional (as illustrated in (8c) and in (10)): it provides information as to the manner, time or place of the event expressed in the sentence:

- 10a Maigret will imitate Poirot.
 10b Bertie will abandon the race.
 10c Miss Marple will reconstruct the crime.

In the traditional literature on parsing, optional phrasal constituents such as the PPs in (8a) and (9) are called *adjuncts*.⁵ While the PPs in the examples above are optional, we see that the VP-internal NPs are not:

- 11a *Maigret will imitate.
 11b *Bertie will abandon.
 11c *Miss Marple will reconstruct.
 11d *Jeeves will meet.

This does not mean, however, that each English sentence contains just one VP-internal NP, as (12) shows:

- 12a Hercule is dithering.
 12b Wooster gave Jeeves the money.

⁵ In chapters 8 and 9 we shall turn to a more technical definition of the notion adjunct, as it is used in the Government and Binding literature.

Inse

13a

On
alter13b
13cIr
mor
clas
intr
obje
verl
obje
NP-
ditr
verl
N
sent
knc
app
knc
mei
nat
C
verl
Thi14a
14b
14c
14c6 J
/

Inserting an NP into the VP of (12a) renders the example ungrammatical:

13a Hercule is dithering *the crime/*Agatha.

On the other hand, in (12b) the verb *give* must be followed by two NPs, or alternatively by an NP and a PP:

13b *Wooster gave Jeeves.

13c Wooster gave [_{NP} the money] [_{PP} to Jeeves].

In traditional grammar the requirement that there should be or not be one or more NPs inside the VP is seen as a property of the verb involved. At least three classes of verbs are traditionally distinguished: transitive, ditransitive and intransitive verbs.⁶ If a VP has a transitive verb as its head, one NP (the direct object) is required: the verb takes an NP-complement. If a VP has a ditransitive verb as its head, two NPs or an NP and a PP (the direct object and the indirect object) are required. If a VP contains an intransitive verb as its head then no NP-complement is allowed. Whether a verb belongs to the group of transitive, ditransitive or intransitive verbs is treated as an idiosyncratic property of the verb.

Native speakers of English would agree on the judgements given for the sentences in (12) and (13). This means that they too must have internal knowledge of the principles that decide on the type of VP in which a verb may appear. The subclassification of verbs must constitute part of their lexical knowledge. Let us therefore integrate the information on verb classes in the mental lexicon which we have posited as part of the internal knowledge of the native speaker.

One way of encoding the information on the complement structure of the verb is by associating it directly with the lexical entry of the verb in question. This would lead to the following (partial) lexicon:

- 14a *meet*: verb; transitive
- 14b *imitate*: verb; transitive
- 14c *reconstruct*: verb; transitive
- 14d *abandon*: verb; transitive

⁶ For discussion of the classification of verbs in the traditional literature, see Aarts and Aarts (1982), Burton-Roberts (1986), Huddleston (1984) and Quirk et al. (1985).

- 14e *dither*: verb; intransitive
 14f *give*: verb; ditransitive

The child acquiring English will have to learn not only the syntactic category of the words of his language, but also the subcategory the verbs belong to. Exposure to English sentences containing these verbs will offer positive evidence of this information: the verb will occur in the appropriate syntactic environment. The child exposed to a sentence like (15a) will thus be able to conclude that *sleep* is intransitive and will assign it the property 'intransitive' in its lexical entry (15b):

- 15a Mummy is sleeping.
 15b *sleep*: verb; intransitive.

In the Chomskian tradition the notions transitive, intransitive, etc. are encoded in distributional frames. Verbs are classified according to the type of VP in which the verb typically occurs. For example, the verb *meet* requires an NP-complement; its VP will contain an NP. This requirement can be represented as follows:

- 16 *meet*: V, [—— NP].

(16) shows in which syntactic frame the verb *meet* can and must be inserted: *meet* is inserted in front of an NP. The verbs are characterized on the basis of the frames in which they occur. *Dither*, for instance, does not take any complement:

- 17 *dither*: V, [——]

The frames that identify subcategories of verbs are called *subcategorization frames*. We say that *meet* subcategorizes for or selects an NP.

3.2 Argument Structure and Thematic Structure

All we have done so far is classify verbs according to whether they require any VP-internal NP. We have not really attempted to explain anything. On the contrary, we have implied that the subcategorization frame of a verb, i.e.

whether
of the
Howev
not a
expres
A ve
active
person
intuiti
the ap
the lo
the ge
thema

3.2.1

Logici
seman
notati

18a
18b

(18a)
expre
things
predi
defin
in (18
m (fo
two-1
appre
predi
is onc
Int
argur

⁷ Fe
(1'

whether it is transitive or intransitive, etc. is an unexplained **primitive** property of the grammar, i.e. a property which does not follow from anything else. However, this seems intuitively wrong. Whether a verb is transitive or not is not a matter of mere chance; it follows from the type of action or state expressed by the verb, from its meaning.

A verb like *imitate* expresses an activity that involves two participants: the active participant, the person who imitates, and the passive participant, the person or thing that is imitated. In Government and Binding Theory this intuitive idea of 'participants in an activity' has been formalized on the basis of the approach commonly adopted in logic. In this section we first look briefly at the logical system of representation, then we apply it to language in terms of the general notion of argument structure and of the more refined notion of thematic structure.

3.2.1 ARGUMENT STRUCTURE IN LOGIC

Logicians have long been concerned with formulating representations for the semantic structure of sentences, or more correctly propositions.⁷ In the notation of formal logic, (18a) is assigned the representation (18b):

18a Maigret imitates Poirot.

18b $A(m, p)$

where A = 'imitate', m = 'Maigret' and p = 'Poirot'.

(18a) contains the NPs *Maigret* and *Poirot*, two referring expressions, i.e. expressions which serve to pick out an entity, a person, a thing, from those things we are talking about, the universe of discourse. It also contains a predicate *imitate*. The predicate does not refer to a person or thing but rather defines some relation between the referring expressions. In the logical notation in (18b) we see that the predicate 'imitate' takes two arguments, represented by m (for Maigret) and p (for Poirot). Predicates that require two arguments are **two-place** predicates. The transitive verbs of traditional syntax correspond approximately to the two-place predicates of logic. The arguments of a predicate are realized by noun phrases in our example: in (18a) the subject NP is one argument and the object NP is the second argument of the verb *imitate*.

Intransitive verbs correspond to **one-place** predicates: they take only one argument.

⁷ For an introduction to logic written specifically for the linguist see Allwood et al. (1977).

19a Maigret stumbled.

19b S (m)

where S = 'stumble' and m = 'Maigret'.

specify
the spe
labels
illustra
ments

3.2.2 ARGUMENT STRUCTURE IN NATURAL LANGUAGE

Using the basic idea of formal logic outlined above, we can say that every predicate has its **argument structure**,⁸ i.e. it is specified for the number of arguments it requires. The arguments are the participants minimally involved in the activity or state expressed by the predicate.

We could use a metaphor to summarize this: predicates are like the script of a play. In a script a number of roles are defined and will have to be assigned to actors. The arguments of a predicate are like the roles defined by the script of a play. For an adequate performance of the play, each role must be assigned to an actor. It will not do either to miss out on a part in the play or to have actors on the stage who have no part to play. Adjuncts might be compared to the parts in the script which are not central to the play.

We first discuss the argument structure of verbs and its relation to subcategorization frames. Then we also turn briefly to the argument structure of adjectives, nouns and prepositions.

The argument structure of the verb determines which elements of the sentence are obligatory. If a verb expresses an activity involving two arguments, there will have to be at least two constituents in the sentence to enable these arguments to be expressed. This conceptually-defined argument structure can partly replace the classification of verbs in terms of either transitivity labels or subcategorization frames described above. If a speaker knows the meaning of the verb *meet*, in other words if he knows what activity is expressed, he will also know how many participants are involved and hence how many arguments the verb takes. 'Meet' involves two participants, and hence will be expected to take two arguments. If one argument is realized as the subject of the sentence (cf. section 6 for discussion), it follows that *meet* will select one VP-internal complement.

This does not mean that we can conclude that the verb *meet* necessarily subcategorizes for an NP. After all, the arguments might have been realized by categories other than NP.⁹ The argument structure of the verb predicts the number of constituents needed but not necessarily their type. Let us assume for the moment that the type of constituent which realizes the argument must be lexically encoded. We can improve the lexical representation of verbs by

20a

20b

20c

20d

20e

20f

20g

Re
conta
with

10 T
str
co
in
ha
dc

ar
se
11 G

(
(i

T.
gi
fc
in

(i

⁸ For a more careful statement see section 5.3 where we discuss the difference between auxiliaries and main verbs.

⁹ We return to this point in sections 5.1 and 7.1.

specifying their argument structure, which is derived from their meaning, and the specification of the realization of the arguments. This notation replaces the labels transitive, intransitive and ditransitive, or the subcategorization frames illustrated in (14) and (16)–(17) respectively.¹⁰ We shall represent the arguments the verb takes by arabic numerals.

20a	<i>meet</i> :	verb;	1	2	
			NP	NP	
20b	<i>imitate</i> :	verb;	1	2	
			NP	NP	
20c	<i>reconstruct</i> :	verb;	1	2	
			NP	NP	
20d	<i>abandon</i> :	verb;	1	2	
			NP	NP	
20e	<i>give</i> :	verb;	1	2	3
			NP	NP	NP
			NP	PP	NP ¹¹
20f	<i>dither</i> :	verb;	1		
			NP		
20g	<i>smile</i> :	verb;	1		
			NP		

Recall that in addition to the arguments of the verb, sentences may also contain adjuncts, constituents providing additional information, for instance with respect to manner, place, time, etc.

¹⁰ There is an important distinction between subcategorization frames and argument structure. Subcategorization frames only specify the complements of the verb, i.e. the constituents that are obligatory inside the VP. The subject NP need not be mentioned in the subcategorization frame because all verbs have subjects, hence the property of having a subject does not create a subcategory of V, whereas the presence of objects does.

The argument structure lists all the arguments: it also includes the subject argument which is realized outside the VP. The thematic structure of the verb (see section 3.2.3) also lists *all* the arguments.

¹¹ *Give* allows for two types of realizations of its arguments:

- (i) I gave Bill the money.
- (ii) I gave the money to Bill.

The representation in (20e) in the text serves to indicate that the second argument of *give* is either realized as an NP (i) or as a PP (ii). As (ii) and (iii) show, a PP must follow the VP-internal NP. We turn to the relative order of VP-internal constituents in chapter 3.

- (iii) *I gave to Bill the money.

In some cases it is less easy to determine the argument structure of predicates. Consider the following pairs of sentences:

21a Hercule bought Jane a detective story.

21b Hercule bought a detective story.

(21a) contains the verb *buy* with apparently three arguments. The argument *Jane* can be omitted, but as a result the meaning of the sentence changes subtly: in the unmarked context (21b) will be taken to mean that Hercule bought the detective story for himself. The action expressed in (21b) still implicitly involves someone for whom the book is bought. (21b) seems to contain an unexpressed or *implicit* argument. We shall encode the fact that some arguments may be left implicit by putting them in parentheses.

22 *Buy*: verb; 1 (2) 3
 NP NP NP

So far we have only illustrated the argument structure of verbs. Other lexical categories too have an argument structure. Consider (23):

23a Poirot is restless.

23b Jeeves is envious of Bertie.

23c Jeeves envies Bertie.

In (23a) the predicate *restless*, an adjective, takes one argument. *Restless* is a one-place predicate. The adjective *envious* in (23b) takes two arguments analogously to the verb *envy* in (23c), which is semantically and morphologically closely related to the adjective. (23b) and (23c) are near-paraphrases. The respective arguments of the verb *envy* in (23c) are realized by the two NPs *Jeeves* and *Bertie*. The arguments of *envious* are realized by an NP and by a PP headed by *of*. At this point we merely note that the second argument of the adjective cannot be realized by a straight NP but that it requires the presence of the preposition *of*. The reason why this should be so is treated in chapter 3.

23d *Jeeves is envious Bertie.

We cannot freely add new referring expressions to the sentences in (23):

24

Unlik

25a

25b

(25b)

again

26a

26b

Th

for ir

27a

27b

We d

is eit

not c

the v

impl

wher

cons

in c

diffe

28a

28b

- 24 *Poirot is restless of the case.

Unlike verbal arguments, the arguments of adjectives can often be left implicit:

- 25a *Poirot envies.
25b Poirot is envious.

(25b) will be understood as meaning 'Poirot is envious of someone.' We shall again encode the argument structure of adjectives in the lexical information:

- 26a *envious*: adjective; 1 (2)
 NP PP
26b *restless*: adjective; 1
 NP

The argument structure of lexical items is not always uniquely fixed. Take for instance the adjective *conscious* in the following examples:

- 27a Miss Marple is conscious of the problem.
27b Sir Galahad is conscious.

We distinguish two argument structures for the adjective *conscious*. *Conscious* is either a two-place predicate (27a) or a one-place predicate (27b). It would not do to say that the second argument of *conscious* is left implicit in (27b) in the way that we argued that the second argument of *envious* in (25b) was implicit. In (25b) the adjective *envious* has the same meaning as in (23b), whereas there is a semantic difference between (27a) and (27b). In (27a) *be conscious* is near-synonymous with *know, be aware*. In (27b) it means 'not be in coma'. Depending on the meaning of the predicate we assume that a different argument structure is associated with it:

- 28a *conscious*₁: adjective; 1 2
 NP PP
28b *conscious*₂: adjective; 1
 NP

*Conscious*₁ will be parallel to *know* or *aware*:

- 29a *know*: verb; 1 2
 NP NP
29b *aware*: adjective; 1 2
 NP PP

Let us turn to nouns. Consider the following groups of examples:

- 30a Poirot will analyse the data.
30b *Poirot will analyse.
30c *There will analyse the data.

- 31a Poirot's analysis of the data was superfluous.
31b The analysis of the data was superfluous.
31c The analysis was superfluous.

In (30) the verb *analyse* requires two arguments. The noun *analysis* is semantically and morphologically related to the verb *analyse* and on the basis of (31a) we assume it has the same argument structure.

- 32a *analyse*: verb; 1 2
 NP NP
32b *analysis*: noun; (1) (2)
 NP PP

The two arguments of *analysis* are realized overtly in (31a); in (31b) the agent of the activity is left unexpressed and in (31c) both arguments are unexpressed. It is a typical property of nouns that both their arguments may be unrealized.

Prepositions too can be argued to have argument structure. The preposition *in*, for instance, will have two arguments; the preposition *between* will have three:

- 33a John is in London.
33b *in*: preposition; 1 2
 NP NP
33c Florence is between Milan and Rome.
33d *between*: preposition; 1 2 3
 NP NP NP

3.2.3

Let us

34a 1

34b 1

In (34

in diff

the su

killing

the ac

argun

numb

but al

to pl

respe

racter

In 1

their

for sh

theta

and 1

argur

the g

theor

Alh

certa

exam

many

types

35a

35b

35c

35d

3.2.3 THETA THEORY

Let us consider the argument structure of the verb *kill*.

34a *Maigret killed Poirot.*

34b *kill*: verb; 1 2
 NP NP

In (34a), the two argument-NPs *Maigret* and *Poirot* are intuitively felt to stand in different semantic relationships with the verb. The argument-NP *Maigret* in the subject position refers to the entity that is the **AGENT** of the activity of killing. The argument NP *Poirot*, the direct object, expresses the **PATIENT** of the activity. We used the metaphor of the script of a play when discussing argument structure of predicates. A script of a play defines not only the number of parts to be assigned, hence the number of actors involved, but also what characters are involved, it specifies which roles these actors have to play. The more specific semantic relationships between the verb and its respective arguments may be compared with the identification of the characters in a play script.

In the literature the more specific semantic relationships between verbs and their arguments are referred to in terms of **thematic roles** or **theta roles** (θ -roles) for short. We say that the verb *kill* takes two arguments to which it assigns a theta role: it assigns the role **AGENT** to the subject argument of the sentence, and the role **PATIENT** to the object argument. The verb **theta-marks** its arguments. Predicates in general have a **thematic structure**. The component of the grammar that regulates the assignment of thematic roles is called **theta theory**.

Although many linguists agree on the importance of thematic structure for certain syntactic processes, the theory of thematic roles is still very sketchy. For example, at the present stage of the theory there is no agreement about how many such specific thematic roles there are and what their labels are. Some types are quite generally distinguished. We discuss them informally here.

- 35a **AGENT/ACTOR**: the one who intentionally initiates the action expressed by the predicate.
- 35b **PATIENT**: the person or thing undergoing the action expressed by the predicate.
- 35c **THEME**: the person or thing moved by the action expressed by the predicate.
- 35d **EXPERIENCER**: the entity that experiences some (psychological) state expressed by the predicate.

- 35e BENEFACTIVE/BENEFICIARY: the entity that benefits from the action expressed by the predicate.
- 35f GOAL: the entity towards which the activity expressed by the predicate is directed.
- 35g SOURCE: the entity from which something is moved as a result of the activity expressed by the predicate.
- 35h LOCATION: the place in which the action or state expressed by the predicate is situated.

The inventory above is very tentative. Other authors amalgamate the roles PATIENT and THEME under the one role of THEME.

- 35i THEME₂: the entity affected by the action or state expressed by the predicate.

We shall usually use the term THEME in this second interpretation.

The thematic roles are illustrated in (36):

- 36a Galahad gave the detective story to Jane
 ACTOR THEME BENEFACTIVE/GOAL
- 36b Constance rolled the ball towards Poirot
 ACTOR THEME GOAL
- 36c The ball rolled towards the pigsty.
 THEME GOAL
- 36d Emsworth sent the parcel to the aged relative.
 ACTOR THEME BENEFACTIVE/GOAL
- 36e Madame Maigret had been cold all day.
 EXPERIENCER
- 36f Madame Maigret was afraid.
 EXPERIENCER
- 36g Poirot bought the book from Maigret.
 ACTOR THEME SOURCE
- 36h Maigret is in London.
 THEME LOCATION

The identification of thematic roles is not always easy, as the reader can verify for himself. However, intuitively the idea should be clear, and we shall be drawing on this rather intuitive approach to theta theory in subsequent discussion.

The
 its argu
 hence
 numbe
 specific
 Bindin
 which
 lexical

37a

(37a)
 We d
 argun
 the sy
 theta

37b

Co

38a

38b

38c

We s
 abse

The information as to the semantic relationship between the predicate and its arguments is part of the lexical knowledge of the native speaker and should hence also be recorded in the lexicon. Rather than merely specifying the number of arguments of a predicate, one may envisage a representation which specifies the type of semantic roles of these arguments. In Government and Binding Theory this is represented by means of a thematic grid, or *theta grid*, which is part of the lexical entry of the predicate. *Kill* would be given the lexical representation in (37a):

37a *kill*: verb

AGENT	PATIENT

(37a) specifies that *kill* assigns two thematic roles (AGENT and PATIENT). We deduce that the verb is a two-place predicate, which requires two arguments to which these roles can be assigned. Some linguists propose that the syntactic category realizing the thematic role should also be specified in the theta grid of a predicate (cf. section 7.1 for discussion).

37b *kill*: verb

AGENT NP	PATIENT NP

Consider some examples:

38a Maigret killed the burglar.

38b *Maigret killed.

38c *Maigret killed the burglar the cellar.

We see that two arguments and no more than two are needed. In (38b) the absence of the second NP renders the sentence ungrammatical: the second

theta role cannot be assigned. In (38c), conversely, one extra NP is added to the sentence. This NP cannot be assigned a thematic role because *kill* only assigns two roles, which are already assigned to the subject-NP and to the object-NP respectively. In (38d) we have inserted the preposition *in*. The sentence is grammatical: the preposition *in* assigns the thematic role of LOCATION to the NP *the cellar*.

38d Maigret killed the burglar in the cellar.

One criterion for judging whether a sentence is grammatical is that the thematic roles associated with its predicate(s) must be assigned to arguments, these arguments must be structurally realized. Conversely, the referring NPs in the sentence must bear some semantic relation to a predicate. This semantic relation can be established via the assignment of thematic roles.

Each syntactic representation of a sentence is scanned for the predicate(s) it contains. Each predicate is tested with respect to its argument structure. Its arguments must be realized. More specifically the predicate is tested for its thematic roles: each role must be assigned to an argument.

Let us take as an example a sentence containing the predicate *kill*. *Kill* assigns the thematic roles of AGENT and PATIENT, hence it requires two arguments. When the theta roles can be assigned to arguments we say that they are saturated and we mark this by checking off the theta role in the thematic grid of the predicate. In order to identify the assignment of the respective thematic roles to the corresponding arguments, NPs are identified by means of an index, a subscript:

- 39a Maigret_i killed the burglar_j.
 39b *Maigret_i killed.
 39c *Maigret_i killed the burglar_j the cellar_k.

We shall not discuss the subscripting convention here. We hope that the intuitive idea is clear: an NP refers to an individual or an object and is identified by the referential index. Two NPs with the same index are said to be co-indexed: they are interpreted as referring to the same entity.¹²

- 40a Maigret_i said that he_i was ill.
 40b Maigret_i hurt himself_i.

¹² We return to a more detailed discussion of co-indexation in chapter 4.

In orde
index c
slot in
represe

41

If we t
grid w
(39c) t
canno

42a

42b

In (4
or no
with
Th
that t
criter

In order to show how the theta roles of a predicate are assigned we enter the index of the argument to which the thematic role is assigned in the appropriate slot in the theta grid. For (39a) the saturation of the thematic roles can be represented as in (41):

41 *kill*: verb

AGENT NP	PATIENT NP
i	j

If we try to do the same for (39b) we see that one of the slots in the thematic grid will remain unfilled: one thematic role is not assigned. Conversely, in (39c) there is one referential index which cannot be entered on the grid, hence cannot be assigned a thematic role.

42a *kill*: verb

AGENT NP	PATIENT NP
i	?

42b *kill*: verb

AGENT NP	PATIENT NP	
i	j	k?

In (42a) corresponding to (39b) the thematic role of PATIENT is not assigned or not saturated. In (42b), corresponding to (39c), the argument NP *the cellar* with the referential index *k* fails to be assigned a thematic role.

The requirement that each thematic role of a predicate must be assigned and that there must be no NPs that lack a thematic role is summed up in the theta criterion:

43 Theta criterion

43a Each argument is assigned one and only one theta role.

43b Each theta role is assigned to one and only one argument.

So far we have only discussed NP arguments. But other constituents may also be arguments: consider, for instance, (44a) and (44b):

44a The police announced the news.

44b The police announced that the pig has been stolen.

In (44a) *announce* is associated with two arguments, which will be assigned their thematic roles. The role AGENT is assigned to *the police*; THEME to *the news*. In (44b) the THEME role is assigned to a subordinate clause: *that the pig has been stolen*. Clauses too can thus be arguments of the predicate. We return to the issue in more detail in section 5.1.

Given the wide diversity among authors on the labelling of thematic roles and their definitions it would be a difficult enterprise to fix the types of roles and their number. Even if we are unable to pin down the exact nature of the different roles involved we are usually quite clear as to how many arguments a predicate requires in a given reading. Hence, instead of specifying the exact type of thematic role for each predicate, we shall often merely list the number of arguments, identifying their roles by numbers rather than by role labels. Thus for the verb *kill* we shall use the following lexical representation, unless we need to refer explicitly to the thematic label.

45 *kill*: verb

1 NP	2 NP

The numerals 1 and 2 represent the thematic roles assigned by the verb whose labels need not concern us.

Recent research¹³ suggests that it might not be necessary or desirable to refer to the thematic labels in the syntax, and that indeed the representation in (45)

¹³ See Belletti and Rizzi (1988). This will not, however, be accessible to the novice.

is the one
reader to t

4 The

Let us sum
which are
syntactic r

Section
the categ
thematic s
minimal c
large exte
principle:

46 Proje
Lexic

The proje

5 Th

In this se
focus on
discusses
lexical ve

5.1 C

We have
the sem:

is the one we need. We do not go into that discussion here and we refer the reader to the literature.

4 The Projection Principle

Let us sum up what we have done so far. We have seen that the lexical items which are the ultimate constituents of a sentence play an important part in its syntactic representation.

Section 2 shows that the lexical category of the head of a phrase determines the category of the phrase. Secondly, we have seen in section 3 that the thematic structure of a predicate, encoded in the theta grid, will determine the minimal components of the sentence. This idea that lexical information to a large extent determines syntactic structure is summed up in the projection principle:

46 Projection principle

Lexical information is syntactically represented.

The projection principle will play an important role throughout this book.

5 The Assignment of Thematic Roles

In this section we look at the assignment of thematic roles in the syntax. We focus on three areas: section 5.1 discusses clausal arguments; section 5.2 discusses expletive pronouns, and section 5.3 considers the difference between lexical verbs, or main verbs, and auxiliaries.

5.1 *Clausal Arguments*

We have seen that the obligatory constituents of a sentence are determined by the semantic properties of the predicates (verbs, adjectives) and we have

mainly discussed examples in which NP-arguments were involved. Sentences too may be arguments of a predicate.

47a Miss Marple has announced the news.

47b Miss Marple has announced that Poirot has left.

In (47a) the verb *announce* takes two arguments, realized by the NPs *Miss Marple* and *the news* respectively. In (47b) the arguments are realized by an NP and by the clause [*that Poirot has left*]. Consider also the following examples:

48a The robbery surprised all the inhabitants of Blandings.

48b [That the pig was stolen] surprised all the inhabitants of Blandings.

49a Jeeves' decision is very unfortunate.

49b [That Jeeves should be leaving] is very unfortunate.

50a Poirot asked three questions.

50b Poirot asked [whether anyone had seen the pig].

51a Maigret believes the story about the burglary.

51b Maigret believes [that the taxi driver is lying].

52a Constance is aware of the problem.

52b Constance is aware [that the pig is in danger].

The verb *surprise* takes two arguments. In (48a) both arguments are realized by NPs; in (48b) one argument is realized by a clause. Similarly, in (49a) the one argument of the adjective *unfortunate* is realized by an NP and in (49b) it is realized by a clause. In (50) and (51) we find further alternations between NPs and clauses as realizations of arguments. In (52) one of the arguments of the adjective *aware* is realized by an NP contained in a PP in (a) and by a clause in (b).

We conclude that the theta grid of predicates will not always specify a unique category to which a theta role can be assigned but will allow for a choice. We do not go into this property in any detail here.

Let us consider clausal arguments a little more closely:

53a [That Galahad had left] is very surprising.

53b [For Galahad to have left] is very surprising.

54a N

54b N

54c N

54d N

In (53

which

(53a):

introdu

predica

senten

princip

In (5

argum

the co

The

form a

essenti

that th

one ar

realize

deduc

55

In (

In (5

(56),

¹⁴ Ko

cla

the

- 54a Maigret_i believes [this story]_j.
 54b Maigret_i believes [that the taxi driver is innocent]_j.
 54c Maigret_i believes [the taxi driver to be innocent]_j.
 54d Maigret_i believes [the taxi driver innocent]_j.

In (53) we see that the adjective *surprising* takes one argument, to which it assigns a thematic role. The argument is realized by a finite clause in (53a): the verb *had* is finite, it is inflected for the past tense and the clause is introduced by the complementizer *that*. In (53b) the argument of the main predicate is realized by a non-finite clause: *have* is in the infinitive and the sentence is introduced by the complementizer *for*. We return to the general principles of sentence structure in chapter 2.¹⁴

In (54a) both arguments of *believe* are realized by NPs. In (54b) one of the arguments of *believe* is realized by a finite clause. As the bracketing indicates, the corresponding argument is realized by a non-finite clause in (54c).

The bracketing in (54c) is meant to show that we consider *the taxi driver* to form a constituent with *to be innocent*. The justification for this analysis is essentially one of analogy. If we compare the sentences (54b) and (54c) we see that they are very similar in meaning. In (54b) the verb takes two arguments: one argument which is realized by the subject NP, and one argument which is realized by a sentence. On the basis of examples like (54a) and (54b) we deduce that the lexical entry of *believe* has the following theta grid:

55 *believe*: verb

1 NP	2 NP/S

In (54a) the arguments are saturated as in (56), where *j* is the index of an NP. In (54b), similarly, the saturation of the arguments can be represented as in (56), with *j* now seen as the index of a subordinate clause.

¹⁴ Koster (1978b) provides arguments that what looks like a clausal subject is not a clause in the subject position. We refer the reader to Koster's own work. However, the article will not be readable until we have worked through chapter 7.

56 *believe*: verb

1 NP	2 NP/S
i	j

Given the close similarity in meaning between (54b) and (54c), the minimal assumption is that the verb *believe* in (54c) is the same as that in (54b) and has the same theta grid. While in (54b) the second argument is associated with a finite clause, in (54c), the second argument is associated with a non-finite clause.¹⁵ Again the theta roles in (54c) are saturated as in (56), with *j* standing for a non-finite clause.

(54d) is also very close in meaning to (54b) and (54c), so we postulate that the verb *believe* is unaltered and has the theta grid in (55). Given this assumption, we need to assign to (54d) a structure that allows the saturation of the argument-roles 1 and 2. The bracketing in (54d) will do that adequately.

It is not immediately obvious how to label the structure [*the taxi driver innocent*]. In the traditional literature on parsing, the term 'verbless clause' is sometimes used. This term serves to indicate that we have a constituent which has a propositional meaning, i.e. the same sort of meaning as a full clausal structure has, but it lacks any verb forms. In (54d) the constituent [*the taxi driver innocent*] corresponds to the sentence [*the taxi driver to be innocent*] in (54c). In both sentences the NP *the taxi driver* is the subject of the predicate expressed by the AP *innocent*. In the Government-Binding literature, constituents such as [*the taxi driver innocent*] are called **small clauses**. We return to their structure throughout the book.

Non-finite clauses and small clauses are not normally¹⁶ found as independent clauses: they can only be subordinate to some other main predicate. The italicised constituents in (57) are all small clauses:

¹⁵ Note, however, that in this particular example, the non-finite clause cannot be introduced by the complementizer *for*. We shall return to this issue in chapters 2 and 3.

¹⁶ Small clauses seem to be in frequent use in certain registers, such as informal notes or telegrams (i) or newspaper headlines (ii):

- (i) Mother in hospital.
- (ii) Hijackers under arrest.

Register-specific syntactic properties have not often been studied in the generative framework (see Haegeman, 1987, 1990; and Massam and Roberge, 1989).

We also find small clauses in colloquial expressions such as:

- (iii) What? Me angry?

- 57a I consider *John a real idiot*.
 57b The chief inspector wants *Maigret in his office*.
 57c Emsworth got *Galahad in trouble*.

It is evident that the small clauses are of different types. In (57a) the small clause consists of an NP *John* and a second NP *a real idiot*. The first NP acts as a subject to the second one. In (57b) and (57c) the small clause is composed of an NP and a PP, where the NP is the subject with respect to the PP predicate.¹⁷ That the italicized strings in (57) are constituents is supported by the fact that other material associated with the main verb of the sentence cannot occur internally to what we have called the small clause:

- 58 *The chief inspector wants [Maigret [very much] in his office].



In (58) the degree adjunct *very much*, which modifies the verb *want*, cannot intervene between the subject and the PP predicate of the small clause.

5.2 Expletives

Section 5.1 shows that not all arguments of a predicate are necessarily realized as NPs. In this section we shall see that the reverse also holds: some NPs in the subject position of the sentence are not assigned a thematic role, hence are not arguments.

5.2.1 IT AND EXTRAPOSITION

The obligatory presence of certain constituents in a sentence can be accounted for in terms of the argument structure of the predicate of a sentence. Let us now extend our analysis to some further data:

- 59a The burglary surprised Jeeves.
 59b That the pig had been stolen surprised Jeeves.
 59c It surprised Jeeves that the pig had been stolen.

¹⁷ For a discussion and further motivation of the analysis of small clauses, see Stowell (1983). However, this work will only become accessible once chapter 8 has been covered.

From (59a) and (59b) we deduce that *surprise* takes two arguments. Neither of these can be omitted:

- 60a *The burglary surprised.
 60b *Surprised Jeeves.
 60c *That the pig had been stolen surprised.

Surprise will be associated with the thematic grid (61):

61 *surprise*: verb

1 NP/S	2 NP

We cannot insert another NP in these sentences since this would not be able to be assigned argument status as it would not receive a theta role from *surprise*.

- 62a *The burglary surprised Jeeves *it*.
 62b *That the pig had been stolen surprised Jeeves *it*.

In (62a) or (62b) the NP *it* cannot be assigned a thematic role and thus the sentence violates the theta criterion (43). The theta criterion specifies that theta roles are assigned uniquely. Hence one could not, for instance, propose that in (62a) theta role 1 is assigned both to the subject NP *the burglary* and to the NP *it*.

Now let us look at (59c) repeated here as (63a):

- 63a It surprised Jeeves that the pig had been stolen.

(63a) is a paraphrase of (59b). We deduce that *surprise* in (63a) has the theta grid given in (61) with two theta roles to be associated with two arguments. How are these arguments realized? If we capitalize on the equivalence between (63a) and (59b) then the easiest thing would be to say that in both (59b) and (63a) one theta role, say 1, is assigned to the clause [*that the pig had been stolen*] and the other one, 2, to *Jeeves*. This hypothesis leaves us with the

NP-constituent *it* in the subject position of (63a) unaccounted for. This NP is not optional:

63b *Surprised Jeeves that the pig had been stolen.

On the other hand, *it* cannot be assigned a thematic role since *surprise* only assigns two thematic roles already saturated as described above.

One element in the discussion is that the choice of a filler for the subject position in (63a) is very limited: indeed no other NP (pronominal or not) can fill the position:

64a **This* surprised Jeeves that the pig had been stolen.

64b **He* surprised Jeeves that the pig had been stolen.

Moreover it is not possible to question the element *it* in (63a):

64c **What* surprised Jeeves that the pig had been stolen?

In fact, the pronoun *it* in (63a) contributes nothing to the meaning of the sentence, (63a) being a paraphrase of (59b). *It* is not a referring expression: it does not refer to an entity in the world, a person or an object; it cannot be questioned.

On the basis of these observations we formulate the hypothesis that *it* plays no role in the semantic make-up of the sentence and that its presence is required in (63a) simply for some structural reason. The relevant explanation for the presence of *it* in the subject position in (63a) will be shown to be that English sentences must have an overt subject (see section 6 and chapter 2 for more discussion). We propose that the pronoun *it* in (63a) acts as a mere slot-filler, a dummy pronoun without semantic contribution to the sentence; it is a place-holder for the otherwise unfilled subject position.

In the literature such a dummy pronoun is often called an **expletive** pronoun. The term **pleonastic** *it* is also used. Expletives are elements in NP positions which are not arguments and to which no theta role is assigned. Note that, unlike adjuncts, expletives contribute nothing to the sentence meaning.

In an example like (63a) it is sometimes said that the sentential subject is **extraposed** and that it is in construction with an expletive. (65) contains some more examples of extraposition patterns. The extraposed clause and the expletive are italicized:

- 65a *It* worries Maigret *that Poirot should have left*.
 65b *It* is unfortunate *that Poirot should have said that*.
 65c *It* is out of the question *that Jeeves should be fired*.
 65d I consider *it* odd *that Poirot should have left*.

The expletive *it* cannot just appear in any type of sentence. Consider for instance the following pair:

- 66a An announcement about the robbery worried Maigret.
 66b **It* worried Maigret an announcement about the robbery.

(66b) shows that the expletive *it* cannot become the place-holder for an extraposed NP.¹⁸

5.2.2 THERE AND EXISTENTIAL SENTENCES

Now let us turn to another sentence pattern which poses problems for our theory outlined so far.

- 67a Three pigs are escaping.
 67b There are three pigs escaping.

In (67a) the predicate 'escape' has one argument, realized by the NP *three pigs*. In (67b) the sentence contains one more element: the pronominal *there*, which occupies the subject position. First note that *there* is not an adjunct of place. In (67b) *there* cannot be questioned like other place adjuncts by means of *where*:

- 68a I saw Bill there last week.
 Where did you see Bill last week?
 There.
 68b *Where are three pigs escaping? There.

Also, unlike the place adjunct, *there* in declarative (67b) cannot be omitted freely:

¹⁸ A good survey of the literature on extraposition can be found in Williams (1980). For different views see also Bennis (1986), Grange and Haegeman (1989) and Postal and Pullum (1988).

69a
69b

But *th*
has *th*
struct
pronc
there
sender

69c

The
there-
are a
const

70a
70b

5.2.3

We h
used
subje
prese
Exj
positi
so far
Exple
rizes:
requi
the N
becau
(71)

¹⁹ Fo
19'
cov

69a I saw Bill last week.

69b *Are three pigs escaping.

But *there* does not really contribute anything to the meaning of (67b), which has the same meaning as (67a). Again the data suggest that *there* is required for structural reasons: it fills up the subject position. As was the case with the pronominal *it* discussed before, we call *there* an expletive. In contrast with *it*, *there* is used to replace NP-subjects which have been moved to the right in the sentence, and it cannot replace clausal subjects:

69c *There surprised Jeeves [that the pig had been stolen].

The construction with *there* has many intriguing properties. For instance the *there*-construction is only allowed if the moved subject NP is indefinite. There are also heavy restrictions on the type of verb that can occur in this construction. Transitive verbs, for instance, are disallowed.¹⁹

70a *There are the three pigs escaping.

70b *There saw three children the pigs.

5.2.3 CONCLUSION

We have seen that there are two pronouns in English, *it* and *there*, that can be used without being assigned a thematic role. They are expletives filling the subject position for structural reasons. We turn to those structural reasons presently.

Expletives always turn up in the subject position of the sentence, i.e. the NP position for which the verb does not subcategorize. Indeed the theory outlined so far predicts that expletives will never turn up in subcategorized positions. Expletives are elements lacking a theta role. The positions a verb subcategorizes for are determined by the thematic structure of the verb. Whenever a verb requires a complement NP, this is because the verb has a theta role to assign to the NP. Inserting an expletive NP in an object position would miss the point, because the expletive element would not be able to receive the theta role. In (71) we find a pronoun *it* as the object of *believe*, but this pronoun is not an

¹⁹ For the discussion of the *there*-construction, see Belletti (1988), Milsark (1974, 1977) and Stowell (1978). These texts will be accessible after chapter 6 has been covered.

expletive: it is assigned a thematic role by the verb. In such examples *it* can substitute for other NPs:

71 Poirot believes *it*/this/this story/the announcement.

The prediction of the theory outlined is thus that expletives can only occur in NP positions that are not subcategorized for, i.e. the subject position of the sentence.²⁰

5.3 Main Verbs and Auxiliaries

So far we have implied that all verbs assign thematic roles. However, it is well known that the class of verbs can be divided into two sets: (a) lexical verbs or main verbs like *eat*, *sleep*, *walk*, and (b) auxiliaries: *be*, *have*, *do*, and the modal auxiliaries *will*, *shall*, *can*, *may*, *must*, *ought*. All these elements are inflected for tense:²¹

72	Verb	Present tense		Past tense
a	<i>eat</i>	eat	eats	ate
b	<i>sleep</i>	sleep	sleeps	slept
c	<i>walk</i>	walk	walks	walked
d	<i>be</i>	am/are	is	was/were
e	<i>have</i>	have	has	had
f	<i>can</i>	can	can	could
g	<i>do</i>	do	does	did

Auxiliary verbs have some special properties distinguishing them from lexical verbs. In (73) and (74) we have paired sentences containing a lexical verb in (a) and an auxiliary in (b). The reader can check that auxiliaries and main verbs behave differently in negative and interrogative patterns:

73a John eats chocolate.
 *John eatn't chocolate.
 John doesn't eat chocolate.

²⁰ See Postal and Pullum (1988) for a different view.

²¹ In the present tense, verbs and the auxiliaries *have* and *be* are also inflected for person and number. Modals are not inflected. For a history of the development of modals see Lightfoot (1979).

- 73b John has eaten chocolate.
 John hasn't eaten chocolate.
 *John doesn't have eaten chocolate.

- 74a John eats chocolate.
 *Eats John chocolate?
 Does John eat chocolate?
 74b John has eaten chocolate.
 Has John eaten chocolate?
 *Does John have eaten chocolate?

The negation element *n't* follows the auxiliaries (cf. (73b)), whereas it must precede the lexical verb (cf. 73a)). In a *yes-no* question the auxiliary and the subject of the sentence are inverted (see chapter 2 for discussion). Lexical verbs do not invert with their subjects: in both negative sentences and in questions the auxiliary *do* is needed. Now let us consider the thematic structure of auxiliaries and main verbs.

- 75a Poirot accuses Maigret
 75b Poirot has accused Maigret.
 75c Poirot is accusing Maigret.
 75d Poirot does not accuse Maigret.

In (75a) the assignment of the thematic roles of *accuse* is straightforward: one thematic role will be assigned to the NP *Poirot* and the other one to *Maigret*. In addition to the lexical verb *accuse*, (75b) contains the perfective auxiliary *have*. The sentence is grammatical, which must mean that all thematic roles of the predicate(s) are assigned and that all referring NPs in the sentence have a thematic role assigned to them. Given that there are just as many NPs present in (75b) as in (75a), we are led to conclude that the auxiliary *have*, though morphologically like a verb in that it is inflected for tense, person and number, does not assign any thematic roles of its own. If *have* did assign any thematic roles then we would expect (75b) to contain one or more NPs in addition to those in (75a), which would be assigned the thematic roles of the auxiliary. The same argument can be applied to the auxiliaries *be* in (75c) and *do* in (75d). We conclude that auxiliaries do not assign thematic roles.

A related problem appears in connection with the copula *be* in (76) (cf. example (54)).

- 76a Maigret_i believes [that the taxi driver is innocent]_i

76b Maigret_i believes [the taxi driver to be innocent]_i

76c Maigret_i believes [the taxi driver innocent]_i

In (76) *believe* assigns one theta role to *Maigret* and it assigns the second one to the bracketed clausal constituents. We are concerned with the internal predicate argument structure of the clausal argument. In the discussion above we have argued that the finite complement clause in (76a), the non-finite one in (76b) and the small clause in (76c) all basically mean the same thing: in all of them the property *innocent* is ascribed to the referent of the NP *the taxi driver*. We have also seen that adjectives, like verbs, have an argument structure. Let us first turn to (76c). Inside the small clause *the taxi driver innocent* the NP *the taxi driver* must have been assigned a thematic role, by virtue of clause (43a) of the theta criterion. We deduce that the NP is assigned a thematic role by *innocent*. The adjective will have the thematic grid (77):

77 *innocent*: adjective

1

In the small clause in (76c) the theta role is assigned to the NP *the taxi driver*. With respect to (76a) and (76b) *innocent* must also be able to assign its thematic role and by analogy with (76c) we assume that it assigns it to the NP *the taxi driver* and hence we conclude that the copula *be*, like auxiliaries, does not assign any thematic roles. Interestingly, the copula *be* also has the other syntactic properties of the auxiliaries:

78a The taxi driver is innocent.

78b The taxi driver isn't innocent.

78c Is the taxi driver innocent?

The formal differences between main verbs on the one hand and auxiliaries and the copula *be* on the other are matched by a semantic property: neither auxiliaries nor the copula *be* assign thematic roles.²²

²² See Pollock (1989) for an explanation. This article is very advanced and should not be tackled until the whole of this book has been covered.

6 The Extended Projection Principle (EPP)

Our discussion reveals that sentence constituents may be required for two reasons.

In the first place, the argument structure and the theta grid of the predicate determine the minimal composition of the sentence. Sentence structure is thus partly lexically determined. This property of syntactic representations is summed up in the projection principle ((46), section 4).

Secondly, expletive elements are required to fill the subject position in certain constructions (section 5.2). The structural requirement which necessitates the insertion of expletives is that sentences must have subjects.²³ This requirement is not one that is specific to individual lexical items, but it is a general grammatical property of all sentences. In this respect the structural requirement that sentences have subjects is an addition to the projection principle. Not only must lexical properties of words be projected in the syntax, but in addition, regardless of their argument structure, sentences must have subjects. The latter requirement has come to be known as the **extended projection principle (EPP)** (79). The phrase structure rules of our grammar (cf. (2b)) will specify that every sentence has a subject. (We return to a discussion of phrase structure in chapter 2.)

79 Extended projection principle

$S \longrightarrow NP - AUX - VP$

Consider (80):

80a *Maigret* accused *Poirot*.

80b *Accused *Poirot* *Maigret*.

In both (80a) and (80b) the two arguments of *accuse* are realized by the NPs *Maigret* and *Poirot*. The ungrammaticality of (80b) follows from the extended projection principle (79): the subject position is not filled. Insertion of the expletive *there* is not possible because *there* cannot be associated with definite NPs and also it cannot be used with transitive verbs. Similarly, *it* cannot be inserted since this expletive cannot be in construction with an NP.

²³ In chapters 5–8 we shall see that the subject may be non-overt.

81a *There accused Maigret Poirot

81b *It accused Maigret Poirot.

7 Thematic Roles: Further Discussion

7.1 The Syntactic Realization of Theta Roles

Linguists do not agree about the extent to which the syntactic category with which a thematic role is associated must be signalled in the thematic grid. It has been noted, for instance that the role ACTOR is always realized as an NP. NP could be said to be the canonical realization of ACTOR. Research is in progress to determine whether such canonical realizations can be generalized and what type of exceptions arise. In our discussion we shall often omit the specification of the syntactic category to which a thematic role is assigned and use representations such as (82). This convention is adopted for convenience' sake and implies no decision with respect to the issue discussed here.

82 *kill*: verb

1	2

7.2 The Subject Theta Role

With respect to the assignment of thematic roles we have treated arguments in the subject position in the same way as arguments in the object position. For instance in (83) we would have said that both the NP *Maigret* and the NP *the taxi driver* are assigned a theta role by the verb *accuse*.

83 Maigret accused the taxi driver.

Ty
subje
choic
the s
hanc
are r
In
the s

84a
84b

In b
the
John
whc
mea
free
corr

85a
85b
85c
85c
(ex.

7
det
the
ass
las'
I
say
the

Two related observations are often advanced in the literature for treating subject arguments as different from object arguments. On the one hand, the choice of the object affects the thematic role of the subject while the choice of the subject argument does not affect the role of the object, and on the other hand, there exist 'object idioms' with the subject as a free argument while there are no subject idioms with a free object.²⁴

In (84) we see how the choice of the object may determine the theta role of the subject:

- 84a John broke a leg last week.
- 84b John broke a vase last week.

In both (84a) and (84b) the verb *break* takes an NP-complement. The choice of the complement determines the thematic role of the subject: while in (84b) *John* could be considered ACTOR, in (84a) this is not the case: John is the one who undergoes the event. Consider also (85). While in (85a) the literal meaning of the verb *kill* is intended, the other three examples are idioms with free subjects. The idiomatic interpretation of the sentence depends on the combination of the verb *kill* and its object.

- 85a kill an insect
 - 85b kill a conversation
 - 85c kill a bottle (i.e. empty it)
 - 85d kill an audience (i.e. wow them)
- (examples from Bresnan, 1982: 350).

The theta role assigned to the subject is assigned **compositionally**: it is determined by the semantics of the verb and other VP constituents. Roughly, the verb assigns an object role first, the resulting verb-argument complex will assign a theta role to the subject. The subject argument is as it were slotted in last.

If a predicate assigns a thematic role directly to some constituent we shall say that the predicate **theta-marks** the constituent **directly**. If the predicate theta-marks an argument compositionally we call this **indirect theta-marking**.

²⁴ A discussion of the grammatical functions in the Government and Binding framework is found in Marantz (1981). For a discussion of some problems raised by the approach, see Bresnan (1982).

As mentioned above there is no agreement as to whether the difference between the types of theta roles should be considered as syntactically relevant. However, most linguists agree that the thematic role assigned to the subject must somehow be set apart from the other thematic roles. One quite popular proposal due to Edwin Williams (1981)²⁵ is that the argument which must be realized in the subject position and hence will be theta-marked indirectly is singled out lexically. The lexical entry for the predicate signals explicitly which argument must be outside the VP. Given that this argument is projected onto an NP outside the VP, it is referred to as the **external argument** and conventionally the external argument is indicated in the thematic grid by underlining:

86 *accuse*: verb

<u>1</u>	2

Theta roles assigned to internal arguments will be referred to as **internal** theta roles; that assigned (indirectly) to the external argument is often referred to as the **external** theta role. (Cf. chapter 6, sections 3 and 5 for further discussion.)

8 Summary

In this chapter we have considered the extent to which sentence structure is determined by lexical properties. As a basis for the formation of sentences we have adopted the projection principle:

²⁵ Williams (1981) shows the relevance of the distinction between external and internal theta role for the domain of morphology. See also exercise 6.

In chapter 6 (section 5) we shall see the importance of setting off internal theta roles from external ones. We shall also discuss an alternative approach to the status of the subject-NP.

1 Proj
Lexi

The ty
in this ch
types of
ciated wi
criterion:

2 The
2a Eac
2b Eac

The tl
assignme
are assoc
Indepe
property
the exter

3 Ex
S -

In orde
subject p
English v

9 Ex

Ex

We
of '
de:
(AC
eac

1 Projection principle

Lexical information is syntactically represented.

The type of lexical information with which we have been mainly concerned in this chapter is the thematic structure of the predicate, i.e. the number and types of arguments which the predicate takes. The thematic structure associated with lexical items must be saturated in the syntax, as stated in the theta criterion:

2 Theta criterion

2a Each argument is assigned one and only one theta role.

2b Each theta role is assigned to one and only one argument.

The theta roles of a predicate are represented in a grid-format. The assignment of thematic roles is registered by means of referential indices which are associated with thematic roles.

Independently of the argument structure of the main predicate, it is a general property of sentences that they must have subjects. This property is stated in the extended projection principle (EPP):

3 Extended projection principle

$S \longrightarrow NP - AUX - VP$

In order to satisfy the EPP, so-called expletives may have to be inserted in the subject position of a sentence. Expletives are pronouns such as *it* and *there* in English which are not assigned a thematic role.